JAZYKOVEDNÝ ČASOPIS

JAZYKOVEDNÝ ÚSTAV ĽUDOVÍTA ŠTÚRA

SLOVENSKEJ AKADÉMIE VIED

2

ROČNÍK 68, 2017

DE GRUYTER OPEN

SAP SLOVAK ACADEMIC PRESS

SUBCATEGORIZATION OF ADVERBIAL MEANINGS BASED ON CORPUS DATA

MARIE MIKULOVÁ – EDUARD BEJČEK – VERONIKA KOLÁŘOVÁ – JARMILA PANEVOVÁ
Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

MIKULOVÁ, Marie – BEJČEK, Eduard – KOLÁŘOVÁ, Veronika – PANEVOVÁ, Jarmila: Subcategorization of Adverbial Meanings Based on Corpus Data. Journal of Linguistics, 2017, Vol. 68, No 2, pp. 268 – 277.

Abstract: We introduce a corpus based description of selected adverbial meanings in Czech sentences. Its basic repertory is one of a long lasting tradition in both scientific and school grammars. However, before the corpus era, researchers had to rely on their own excerption; but nowadays, current syntax has a vast material basis in the form of electronic corpora available. On the case of spatial adverbials, we describe our methodology which we used to acquire a detailed, comprehensive, well-arranged description of meanings of adverbials including a list of formal realizations with examples. Theoretical knowledge stemming from this work will lead into an improval of the annotation of the meanings in the Prague Dependency Treebanks which serve as the corpus sources for our research. The Prague Dependency Treebanks include data manually annotated on the layer of deep syntax and thus provide a large amount of valuable examples on the basis of which the meanings of adverbials can be defined more accurately and subcategorized more precisely. Both theoretical and practical results will subsequently be used in NLP, such as machine translation.

Keywords: adverbial meanings, deep syntax, annotation, treebank

1 INTRODUCTION

The description of adverbial meanings (local, temporal, manner, etc.) has a long lasting tradition and has been covered so far in Czech grammars and syntactic monographs in a varying granularity, with more or less detailed specification of the meanings (e.g. [1], [5], [6], [7], [8], [13], [23], [26], [27]).

However, it is well known that the traditional subclassification of adverbials is not grained enough for NLP tasks. For a deep syntax based machine translation (e.g. [4], [16], [28]), it is assumed that deep syntactic annotation narrows "the distance" between the source and the target language. For a successful transfer of a sentence from one language to another, it is necessary to capture all substantial information about the sentence meaning within the deep syntactic representation. The most important part of this representation is an accurate specification of meanings of particular modifications. On the deep syntactic layer in the Prague Dependency Treebanks (which serve as the corpus sources for our research; see their description in Sect. 2), the units of the sentence, i.e. content words together with their auxiliary words, such as prepositions and conjunctions, are represented by nodes of a tree-shaped graph. The tree reflects the underlying dependency structure of the sentence.

DOI 10.1515/jazcas-2017-0036

The types of the (semantic) dependency relations are represented by the "functor" attribute attached to all nodes. The functors represent relatively general categories. However, from the point of view of machine translation, they are not differentiated enough. For example, all the following modifications *na stole* 'on the table'*, pod stolem* 'under the table'*, za stolem* 'behind the table'*, poblíž stolu* 'near the table', etc. are covered by a single functor with a static meaning "where" (marked as LOC). Each of these modifications expresses the general meaning "where"; however, the introduction of a set of "narrower" meanings ("on the given place", "under the given place", "behind the given place", etc.) makes it possible to reflect the semantic differences among them. Thus, it is obvious that such a differentiation among the partial meanings is needed for a complete meaning of the sentence (and for its translation to another language). The requirement of the splitting one functor into more subtle units (called subfunctors here) occurs not only with spatial or temporal adverbials, but it concerns the other functors, e.g. accompaniment (with/without), regard (with respect to/without respect to), comparison (similarity, difference), etc. Illustrative examples of subfunctors were given in [22]. However, a comprehensive list of fine grained categories has not yet been developed.

To carry out a comprehensive and detailed subcategorization of all adverbial meanings and use it as a basis for creating a complete proposal of subfunctors requires a complex view on the theoretical core of the problem together with constant comparisons of proposed solutions with real data. In this paper, we shortly introduce our corpus sources and on the case of spatial adverbials, we describe our methodology used to fulfil our aims.

2 DATA: PRAGUE DEPENDENCY TREEBANKS

Large corpus sources are inevitable for a comprehensive study of subcategorization of all adverbial meanings. While many Czech corpora has morphological annotation (done automatically), we have to take into account the syntax. Nowadays, several richly syntactically annotated corpora, collectively called Prague Dependency Treebanks (PDTs in the sequel; [9]), have been already developed. These corpora provide a large amount of valuable examples that are used as a basis for the determination of subcategorized meanings of adverbials.

The annotation scenario of PDTs is reflected in several detailed annotation manuals (see [17], [18], and [19]). The main features of the annotation style are:

- well-developed dependency syntax theory which is known as the Functional Generative Description (FGD in the sequel; see [22], [24], [25]),
- interlinked hierarchical layers of standoff annotation,
- deep syntactic layer.

In the years 1996 through 2005, the first **Prague Dependency Treebank**[1] (PDT in the sequel; [11], for the latest version 3.0 see [2]) was designed and built. The data in PDT are composed by articles from the Czech daily newspapers. The slightly modified scenario was then used for the annotation of the Prague Czech-English Dependency Treebank, the Prague Dependency Treebank of Spoken Czech, and PDT-Faust corpus.

[1] http://ufal.mff.cuni.cz/prague-dependency-treebank

In contrast to the anchoring original project of PDT, in these treebanks, the morphological and surface syntactic annotations were done automatically and the manually annotated deep syntactic layer does not contain annotation of information structure and some other special annotations. However, annotation of functors, which we are mainly interested in here, is done manually in all four treebanks.

The **Prague Czech-English Dependency Treebank**[2] version 2.0 (PCEDT in the sequel, see [10]) is a manually parsed Czech-English parallel corpus. The English part consists of the Wall Street Journal section of the Penn Treebank [15]. The Czech part, which is used in our research, was translated from the English source sentence by sentence.

The **Prague Dependency Treebank of Spoken Czech**[3] version 2.0 (PDTSC in the sequel, see [20]) contains slightly moderated testimonies of Holocaust survivors from the Malach project corpus[4] and dialogues (two participants chat over a collection of photographs) recorded for the Companions project.[5]

The **PDT-Faust** is a small treebank containing short segments (very often with vulgar content) translated by the various users on the webpage reverso.net.

	PDT	**PCEDT**	**PDTSC**	**Faust**	**Total**
Tokens	833195	1162072	742257	33772	**2771296**
Sentences	49431	49208	73835	3000	**175474**

Tab. 1. Volume of data in Prague Dependency Treebanks

It is obvious that the Prague Dependency Treebank family provides rich language data for our purpose. Altogether, the treebanks include around 180 000 sentences with their deep syntactic annotation (see Table 1 and 2). Moreover, the PCEDT, PDTSC, and PDT-Faust treebanks will be also extended and corrected by manual annotation on the morphological and surface syntactic layers, and together with PDT, they will become a part of the **Consolidated Prague Dependency Treebanks** release in 2018, which will thus contain four different treebanks of Czech, uniformly annotated using the same scenario, with data coming from text, speech and Internet sources.

3 ANALYSIS OF ADVERBIAL MEANINGS

Our approach following the principles of the FGD is based namely on classification given in Novočeská skladba by Vladimír Šmilauer and Mluvnice současné češtiny 2 by Jarmila Panevová et al. Šmilauer´s classification of modifications ([26], pp. č-334) was used as the basis for constituting the set of functors for FGD as well as for the annotation on the deep syntactic layer in PDTs. The description of modifications given in Mluvnice současné češtiny 2 ([23], pp. 39-100) corresponds to the list of functors used for the annotation scenario applied in the PDTs. Similar detailed analysis of adverbials for English is given in [12].

[2] https://ufal.mff.cuni.cz/pcedt2.0/; https://catalog.ldc.upenn.edu/ldc2012t08

[3] http://ufal.mff.cuni.cz/pdtsc2.0

[4] http://ufal.mff.cuni.cz/cvhm/vha-info.htm

[5] http://www.companions-project.org

The starting point for our research is a subdivision of adverbial meanings into related groups which roughly correspond to the categories described in traditional Czech grammars (spatial, temporal, manner, causal, etc.). Then we gradually analyze one group at a time and generalize individual partial meanings of these modifications. The proposed set of subcategorized meanings is based on the detailed analysis of real examples gained from the PDTs.

Firstly, we study all formal realizations for each functor in the PDTs, i.e. we determine which parts of speech, cases, prepositions, and subordinate conjunctions were used to express the meaning of that particular functor. It means that for each functor we create a list of its formal realizations with a sufficient number of examples.

	LOC	**DIR1**	**DIR2**	**DIR3**
Occurrences	79874	17394	1590	28165
The most frequent forms (in all PDTs)	31531 v+6 17215 adv 13122 na+6 3566 u+2 1396 mezi+7 539 za+7 539 pod+7 461 před+7 393 po+6 350 nad+7	11965 z+2 692 od+2 594 adv 18 ze strany+2 4 zpoza+2 2 zpod+2	496 Instr 393 po+6 327 přes+4 73 adv 33 kolem+2 17 mezi+7 14 okolo+2 13 v+6 13 skrz+4 10 podél+2	9415 do+2 4740 adv 3644 na+4 2254 k+3 233 mezi+4 177 pod+4 170 za+7 106 za+4 90 na+6 61 před+4
The most frequent forms in *written* corpus (PDT)	11894 v+6 3902 na+6 1676 adv 1073 u+2 619 mezi+7 198 před+7 144 za+7 135 pod+7 110 kolem+2 94 v oblasti+2	4362 z+2 202 od+2 75 adv 3 zpoza+2 3 ze strany+2 1 zpod+2	207 Instr 96 přes+4 78 po+6 14 adv 10 mezi+7 3 skrz+4 2 vedle+2 2 nad+7 2 na+6 2 mimo+4	2936 do+2 1101 na+4 765 k+3 439 adv 97 mezi+4 60 pod+4 57 za+7 42 za+4 28 před+4 14 proti+3
The most frequent forms in *spoken* corpus (PDTSC)	14151 adv 7494 v+6 5145 na+6 1628 u+2 284 za+7 209 po+6 196 vedle+2 196 mezi+7 186 pod+7 142 před+7	2812 z+2 470 adv 312 od+2 6 ze strany+2 1 zpod+2	266 po+6 163 přes+4 114 Instr 52 adv 23 kolem+2 13 okolo+2 7 podél+2 5 skrz+4 4 mezi+7 3 podle+2	4002 do+2 3579 adv 1670 na+4 908 k+3 98 za+7 63 na+6 59 pod+4 42 za+4 27 mezi+4 24 po+6

Forms which are *only in written* corpus (PDT)	blízko+3 kol+2 na čele+2 nad+4 na úrovni+2 po boku+2 uvnitř+2 v čele+2 v rámci+4	zpoza+2	před+7 skrze+4 vedle+2	do čela+2 na roveň+2 vůči+3

Tab. 2. Raw frequency of forms of spatial adverbials in PDTs. Usually a preposition plus a case or adverbial phrase (adv) or a direct case (Instr).

The values of functors on the deep syntactic layer of PDTs reflect the semantic distinctions roughly corresponding to the traditional classification of adverbials. However, in the PDT scenario, the repertory of functors is used not only for the modifications dependent on verbs, adjectives and adverbs (i.e. of traditional adverbials, e.g. *Kniha leží na stole.* 'The book is lying on the table.'), but also for the modifications dependent on nouns (e.g. *kniha na stole je černá,* 'book on the table is black'). All modifications (dependent on verbs, adjectives, adverbs and nouns) with particular meanings are objects of our studies.

	LOC (where)	**DIR1 (where from)**	**DIR2 (which way)**	**DIR3 (where to)**
in	*v domě*	*z domu*	*domem*	*do domu*
inside	*uvnitř domu*	*zevnitř domu*	*vnitřkem domu*	*dovnitř domu*
inmiddle	*uprostřed domu*	*zprostřed domu*	*prostředkem domu*	*doprostřed domu*
athead	*na čele domu*	*z čela domu*	-	*do čela domu*
indiff	*po domech*	-	-	*po domech*
intarget	-	-	-	*Střílí po lidech.*
on	*na domě*	*s domu*	*po domě*	*na dům*
above	*nad domem*	*znad domu*	*nad domem*	*nad dům*
below	*pod domem*	*zpod domu*	*pod domem*	*pod dům*
behind	*za domem*	*zpoza domu*	*za domem*	*za dům*
front	*před domem*	*zpřed domu*	*před domem*	*před dům*
frontopp	*naproti domu*	*odnaproti domu*	*naproti domu*	*naproti domu*
near	*u domu*	*od domu*	-	*k domu*
beside	*vedle domu*	-	*vedle domu*	*vedle domu*
alongside	*podél domu*	-	*podél domu*	*podél domu*
around	*kolem domu*	-	*kolem domu*	*kolem domu*
across	*přes dům*	-	*přes dům*	*přes dům*
between	*mezi domy*	-	*mezi domy*	*mezi domy*
among	*mezi domy*	-	*mezi domy*	*mezi domy*
outside	*vně domu*	*zvnějšku domu*	*vně domu*	*vně domu*

Tab. 3. Distribution of subfunctors of four spatial modifications

We are aware that a theoretical description based on the relation of form and function needs a transparent and a systematic treatment reflecting the hierarchy

functor – subfunctor. Since the correspondence between forms and their semantic functions is not one-to-one within a single functor, and not even between the form and the meaning within subfunctors, the determination and systemization of these units is considered to be a part of scientific description of language. The necessity of the subcategorization of the functors is further demonstrated by splitting of spatial modifications into 20 subfunctors (cf. Table 3).

4 SUBCATEGORIZATION OF SPATIAL MEANINGS

The functors for spatial meanings are distinguished according to the question specifying the location as follows (cf. [19, p. 474]):

LOC: where? (static modification, simple localization),

DIR1: where from? (directional modification with the meaning of setting out the starting point),

DIR2: which way? (directional modification; the path rather than starting point or destination),

DIR3: where to? (directional modification with the meaning of approaching a destination).

Subfunctors	Forms	Examples
in	*v+6* *na+6* *u+2*	*V tom údolí byly obrovské plantáže čaje.* (PDTSC)
inside	*uvnitř+2*	*A hle, uvnitř paláce stojí nový palác a nové hradby.* (PDT)
inmiddle	*uprostřed+2* *veprostřed+2* *ve středu+2* *vprostřed+2*	*Táborovou kaplí se stal indiánský stan teepee uprostřed tábora.* (PDT)
athead	*v čele+2* *na čele+2*	*Tank v čele kolony obrněnců se řítí na studenta.* (PDT)
indiff	*po+6*	*Kantoroval po mnoha městech.* (PDT)
on	*na+6* *po+6*	*Poskakoval kolem dokola po jevišti.* (Faust)
above	*nad+7*	*Vyvolal jsem to, nad kamny usušil film, nadělal fotky a večer je přinesl.* (PDTSC)
below	*pod+7*	*My jsme bydleli nahoře a oni bydleli pod námi.* (PDTSC)
behind	*za+7*	*Za hranicemi na mě čekala teta.* (PDTSC)
beside	*vedle+2* *po boku+2*	*Když se manželka oběti vrátila domů, pes pokojně seděl vedle mrtvého těla.* (PDT)
alongside	*dle+2* *podél+2* *podle+2*	*Podle Labe jsou břehy osázené duby.* (PDTSC)
front	*před+7*	*Vejprava fauloval před jabloneckou brankou Krejčíka.* (PDT)
frontopp	*proti+3* *naproti+3* *tváří v tvář+3* *čelem k+3*	*Jak je tam ten dům na fotografii, tak ten byl proti domu, kde jsem tehdy bydlela já.* (PDTSC)

near	*u+2* *při+2* *blízko+2* *blízko+3* *v blízkosti+2* *poblíž+2* *nedaleko+2*	*Na nádvoří odcizil i zaparkovanou Škodu 120, ale vozidlo odstavil nedaleko objektu.* (PDT)
around	*kolem+2* *okolo+2*	*Otevřeně se pokračuje v prodeji drog okolo škol, parků a sídlišť.* (PCEDT)
across	*ob+2* *přes+4*	*Bydlely jsme blízko sebe, přes ulici.* (PDTSC)
between	*mezi+7*	*hodina mezi psem a vlkem* (PDT)
among	*mezi+7*	*Bylo to otevřené, ale já jsem byla mezi posledními.* (PDTSC)
outside	*vně+2* *stranou+2* *mimo+2*	*To musí být strašně těžké být o prázdninách mimo domov.* (Faust)
indomain	*v oblasti+2* *v oboru+2* *na poli+2* *v rámci+2*	*Náklady na zaměstnance stoupají mnohem rychlejším tempem v oblasti zdravotní péče než v jiných odvětvích.* (PCEDT)
inlevel	*na úrovni+2*	*Přesun důležitých pravomocí se nezastaví na úrovni republik.* (PDT)

Tab. 4. Subfunctors, forms and examples for LOC functor

Based on the comprised lists of formal realizations and real examples (acquired from all PDT treebanks; comp. the large amount of acquired material in Table 2), we have proposed subfunctors for each of the four spatial functors. An overall overview of 20 proposed subfunctors is shown in Table 3; a detailed list with forms and examples for one spatial functor (LOC) is given in the Table 4. For the labelling of the subfunctors the preposition prototypical for the given meaning is used instead of a metalanguage signs.

	LOC	**DIR1**	**DIR2**	**DIR3**
on	*Děti běhají po trávníku.*	-	*Pojedeme po náměstí.*	-
indiff	*Vysedávali po náměstích.*	-	-	*Putoval po hradech.*
intarget	-	-	-	*Stříleli po lidech.*

Tab. 5. Functors and subfunctors of *po*+6 form

It is obvious that the boundaries between individual semantic distinctions are not always clear; many ambiguities have to be solved. There is no form – meaning isomorphy, one form is used for expressing more meanings and one meaning can be expressed using various forms. For example, with the form *po*+6 (‘on/along/around’) three different meanings of spatial modifications were distinguished. These meanings are schematically represented in Table 5. Combining the LOC functor with the subfunctor *on* captures a move (an action) which has no target but merely happens on the surface (e.g. *Děti běhají po trávníku/na trávníku.* ‘The children are running on the lawn.’). Combining the DIR2 functor with the subfunctor *on* captures a move on

a surface from somewhere to another place (neither the starting point nor the destination is expressed; e.g. *Pojedeme po náměstí (až ke kostelu*). 'We shall go along the square (till to the church).' The *indiff* subfunctor means that the specified location takes place on several places of the same kind at the same time. It applies to all locations where the action usually/often happens (static LOC; e.g. *Vysedávali po náměstích.* 'They used to sit around squares.'), or places where all individual action heads to (dynamic DIR3; e.g. *Putoval po hradech.* 'He travelled around castles.'). The *po*+6 applied for DIR3 modification conveys a specific meaning with semantically limited group of verbs. The direction is here connected with a live target, a victim, at whom the action (mostly negative) is aimed (e.g. *Stříleli po lidech.* 'They shot at people.'). This meaning is captured by *intarget* subfunctor.

The conditions for the distribution of the forms expressing closely related meanings (such as *Děti běhají po trávníku*. 'The children are running on/along the lawn.' vs. *Děti běhají na trávníku*. 'The children are running on the lawn.'; *Stříleli po lidech.* vs. *Stříleli na lidi.* vs. *Stříleli do lidí*. 'They shot at people.'; *Vysedávali po náměstích.* 'They used to sit on/around squares.' vs. *Vysedávali na náměstích.* 'They used to sit on squares.') as well as the cases of lexicalization where two different prepositions express the same meaning (e.g. *Bydlí v Praze/v Dejvicích* 'He lives in Prague/in Dejvice' vs. *Bydlí na Kladně / na Letné*. 'He lives in (lit. 'on') Kladno/on Letná.') are studied. A study of the overlapping of meanings can contribute to the introduction of the new subfunctors. Our goal is to describe and analyse the cases of overlapping meanings from the theoretical point of view as well as in the form of practical guidelines for annotation procedure. Reliable criteria ensuing from the language system itself will be formulated in order to specify the partial meanings and subtle semantic distinctions.

The secondary prepositions and their specific meanings are studied as well as a wide range of expressions which more or less correspond with expressions generally perceived as secondary prepositions. They are temporarily tagged as potential candidates for the word-class of prepositions. For the LOC functor, there are, e.g., the following secondary prepositions: *ve středu*+2 'in the centre of', *v čele*+2 'at the head of', *tváří v tvář*+3 'face to face to', *v oblasti*+2 'in the domain of', *na poli*+2 'in the field of' (see Table 4). The study of criteria for determination of the class of secondary prepositions in Czech and for their semantic and/or stylistic contribution to the meaning of the sentence with regard to the examples from corpora as well as to the results proposed in the printed papers and monographs (e.g. [3], [14]) is needed and it will be presented elsewhere.

5 EXPRESSING OF ADVERBIAL MEANINGS IN WRITTEN AND SPOKEN CZECH

The fact that we currently have different types of annotated corpora of the Czech language, particularly written texts corpus PDT and spoken texts corpus PDTSC offers a unique opportunity to compare expressions of adverbial meanings in written and spoken Czech in a precise and reliable way. The repertory of adverbial meanings and their formal realizations in both types of data has to be compared in more detail.

We expect a refinement of forms for expressing adverbial meanings in written text on the one hand, and marked, peculiar forms in spontaneous speech on the other hand (cf. similar observations in PDT corpora for valency modifications in [21]). Likewise, in a general and simple overview in the Table 2, we can observe that secondary prepositions for abstract and refined meaning (cf. *v oblasti+2* 'in the domain of', *v rámci*+2 'within the frame of', *na úrovni*+2 'at the level of', *po boku*+2 'alongside with', *na čele*+2 'at the head of', *do čela*+2 'to the head of') are more typical for written text. The secondary prepositions occur among the forms which are present only in written corpus and do not occur in spoken one.

6 CONCLUSION

We introduced here our research focused on a description of selected adverbial meanings in Czech sentences. On the case of spatial adverbials, we described our methodology and demonstrated that the Prague Dependency Treebanks provide us with valuable and rich material allowing us to elaborate the issue in depth. We believe that a systematic and accurate description of adverbial meanings verified on the basis of corpus material is necessary for comparative studies and for an application in NLP tasks as well as for a comprehensive syntactic description.

ACKNOWLEDGEMENTS

The research reported in the paper has been supported by the Czech Science Foundation under the projects GA16-02196S and GA17-12624S. The research has also been supported by the LINDAT/CLARIN project of Ministry of Education, Youth and Sports of the Czech Republic (project LM2015071).

References

[1] Bauer, J. and Grepl, M. (1972). *Skladba spisovné češtiny*. SPN, Praha.

[2] Bejček, E., Hajičová, E., Hajič, J., Jínová, P., Kettnerová, V., Kolářová, V., Mikulová, M., Mírovský, J., Nedoluzhko, A., Panevová, J., Poláková, L., Ševčíková, M., Štěpánek, J., and Zikánová, Š. (2013). Prague Dependency Treebank 3.0. Data/software, MFF, ÚFAL, Prague.

[3] Blatná, R. (2006). *Víceslovné předložky v současné češtině*. Lidové noviny, Praha.

[4] Bojar, O., Callison-Burch, Ch., Hajič, J., and Koehn, P. (2009). Special Issue on Open Source Machine Translation Tools. *The Prague Bulletin of Mathematical Linguistics*, 91.

[5] Daneš, F. et al. (1987). *Mluvnice češtiny 3*. Academia, Praha.

[6] Grepl, M. et al. (1997). *Příruční mluvnice češtiny*. Druhé opravené vydání. Lidové noviny, Praha.

[7] Grepl, M. and Karlík, P. (1986). *Skladba spisovné češtiny*. SPN, Praha.

[8] Grepl, M. and Karlík, P. (1998). *Skladba češtiny*. Votobia, Olomouc.

[9] Hajič, J., Hajičová, E., Mikulová, M., and Mírovský, J. (2017). Prague Dependency Treebank. In *Handbook on Linguistic Annotation. Volume II*, pages 555–594, Springer Science+Business Media, Dordrecht, Netherlands.

[10] Hajič, J., Hajičová, E., Panevová, J., Sgall, P., Bojar, O., Cinková, S., Fučíková, E., Mikulová, M., Pajas, P., Popelka, J., Semecký, J., Šindlerová, J., Štěpánek, J., Toman, J., Urešová, Z., Žabokrtský, Z. (2012). Announcing Prague Czech-English Dependency Treebank 2.0. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 3153–3160, European Language Resources Association, Istanbul, Turkey.

[11] Hajič, J., Panevová, J., Hajičová, E., Sgall, P., Pajas, P., Štěpánek, J., Havelka, J., Mikulová, M., Žabokrtský, Z., Ševčíková-Razímová, M., and Urešová, Z. (2006). Prague Dependency Treebank 2.0. Data/Software, Linguistic Data Consortium, Philadelphia.
[12] Hasselgård, H. (2010). *Adjunct adverbials in English*. Cambridge University, Cambridge.
[13] Havránek, B. and Jedlička, A. (1960). *Česká mluvnice*. SPN, Praha.
[14] Kroupová, L. (1985). *Sekundární předložky v současné češtině*. ÚJČ ČSAV, Praha.
[15] Marcus, M., Santorini, B., and Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: the Penn Treebank, *Computational Linguistics*, 19(2):313–330.
[16] Mareček, D., Popel, M., and Žabokrtský, Z. (2010). Maximum Entropy Translation Model in Dependency-Based MT Framework. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 201–202, Association for Computational Linguistics, Uppsala, Sweden.
[17] Mikulová, M. (2014). Annotation on the tectogrammatical level. Additions to annotation manual (with respect to PDTSC and PCEDT). Technical report no. 2014/ÚFAL TR-2013-52, ÚFAL MFF UK, Prague.
[18] Mikulová, M., Bejček, E., Mírovský, J., Nedoluzhko, A., Panevová, J., Poláková, L., Straňák, P., Ševčíková, M., and Žabokrtský, Z. (2013). From PDT 2.0 to PDT 3.0 (Modifications and Complements). Technical report no. 2013/ÚFAL TR-2013-54, ÚFAL MFF UK, Prague.
[19] Mikulová, M., Bémová, A., Hajič, J., Hajičová, E., Havelka, J., Kolářová, V., Kučová, L., Lopatková, M., Pajas, P., Panevová, J., Razímová, M., Sgall, P., Štěpánek, J., Urešová, Z., Veselá, K., and Žabokrtský, Z. (2006). Annotation on the tectogrammatical level in the Prague Dependency Treebank. Annotation manual. Technical report no. 2006/30, ÚFAL MFF UK, Prague.
[20] Mikulová, M., Mírovský, J., Nedoluzhko, A., Pajas, P., Štěpánek, J., and Hajič, J. (2017, in press). PDTSC 2.0 – Spoken Corpus with Rich Multi-layer Structural Annotation. In *Lecture Notes in Computer Science*, Springer, Dordrecht, Netherlands.
[21] Mikulová, M., Štěpánek, J., and Urešová, Z. (2013). Liší se mluvené a psané texty ve valenci? *Korpus – gramatika – axiologie*, 8:36–46.
[22] Panevová, J. (1980). *Formy a funkce ve stavbě české věty*. Academia, Praha.
[23] Panevová, J., Hajičová, E., Kettnerová, V., Lopatková, M., Mikulová, M., and Ševčíková, M. (2014). *Mluvnice současné češtiny 2. Syntax na základě anotovaného korpusu*. Karolinum, Praha.
[24] Sgall, P. (1967). *Generativní popis jazyka a česká deklinace*. Academia, Praha.
[25] Sgall, P. et al. (1986). *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. D. Reidel Publishing Company, Dordrecht.
[26] Šmilauer, V. (1969). *Novočeská skladba*. 2. vydání. SPN, Praha.
[27] Štícha, F. et al. (2013). *Akademická gramatika spisovné češtiny*. Academia, Praha.
[28] Tamchyna, A., Popel, M., Rosa, R., and Bojar, O. (2014). CUNI in WMT14: Chimera Still Awaits Bellerophon. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pages 195–200, Association for Computational Linguistics, Baltimore.